TITLE: Method for stable gene-amplification in a bacterial host cell

FIELD OF INVENTION

In the biotech industry it is desirable to construct polypeptide production strains
having several copies of a gene of interest stably chromosomally integrated, without leaving
antibiotic resistance marker genes in the strains.

This invention relates to bacterial host cells comprising at least two copies of an
amplification unit in its genome, said amplification unit comprising: i) at least one copy of a
gene of interest, and ii) an expressible conditionally essential gene, wherein the conditionally
essential gene is either promoterless or transcribed from a heterologous promoter having an
activity substantially lower than the endogenous promoter of said conditionally essential
gene, and wherein the conditionally essential gene if not functional would render the cell
auxotrophic for at least one specific substance or unable to utilize one or more specific sole
carbon source; methods for producing a protein using the cell of the invention, and methods
for constructing the cell of the invention.

BACKGROUND OF THE INVENTION

In the industrial production of polypeptides it is of interest to achieve a product yield
as high as possible. One way to increase the yield is to increase the copy number of a gene
encoding a polypeptide of interest. This can be done by placing the gene on a high copy
number plasmid, however plasmids are unstable and are often lost from the host cells if
there is no selective pressure during the cultivation of the host cells. Another way to increase
the copy number of the gene of interest is to integrate it into the host cell chromosome in
multiple copies.

The present day public debate concerning the industrial use of recombinant DNA
technology has raised some questions and concerns about the use of antibiotic resistance
marker genes. Antibiotic marker genes are traditionally used as a means to select for strains
carrying multiple copies of both the marker genes and an accompanying expression
cassette coding for a polypeptide of industrial interest. In order to comply with the current
demand for recombinant production host strains devoid of antibiotic markers, we have
looked for possible alternatives to the present technology that will allow substitution of the
antibiotic markers we use today with non-antibiotic marker genes.

WO 02/00907 (Novozymes, Denmark) discloses a method for stable chromosomal
multi-copy integration of genes into a production host cell in specific well-defined sites. It is
disclosed to first render a recipient cell deficient by inactivating one or more conditionally
essential gene, e.g., to make the cell auxotrophic for an amino acid. A gene of interest may
then be integrated into the chromosome along with a DNA sequence which complements

the deficiency of the cell, thus making the resulting cell selectable; the *Bacillus licheniformis metC* gene is disclosed as a conditionally essential marker herein.

WO 01/90393 (Novozymes, Denmark) discloses a method for increasing the gene copy number in a host cell by gene-amplification, without leaving antibiotic resistance markers behind in the host cell. The disclosed method relies on rendering a specific type of conditionally essential chromosomal gene of the host cell non-functional. A single amplification unit comprising the gene of interest, and a DNA sequence, which when integrated into the chromosome complements the non-functional conditional essential chromosomal gene, is integrated into the chromosome.

In order to provide recombinant production strains devoid of antibiotic resistance markers, it remains of industrial interest to find new methods to stably integrate genes in multiple copies into host cell chromosomes. Even incremental improvements of existing methods or mere alternatives are of considerable interest to the industry.

SUMMARY OF THE INVENTION

The problem to be solved by the present invention is to provide alternative host cells comprising multiple copies of a gene of interest, which cells are devoid of antibiotic markers, for use in the industrial production of polypeptides in high yields.

The solution is based on the observation that an amplification unit can be integrated into the chromosome of a host cell, and subsequently be amplified, without the use of classical antibiotic markers, antibiotics, or endogenously produced inhibitory compounds.

In traditional amplification protocols, higher gene expression is a result of duplications of the antibiotic resistance marker gene, duplications which are selected in stepwise cultivation and selection rounds by adding increasing amounts of the antibiotic compound to the cultivation medium in each cultivation step.

A cell which has become auxotrophic, e.g., due to a non-functional conditionally essential gene, would normally be complemented back to the prototrophic phenotype by the integration (or restoration) in the chromosome of even one single functional copy of the non-functional gene. Since normally only one copy is needed, such genes have not previously been attractive candidates for amplification purposes.

However, the present inventors lowered the expression-level of a non-antibiotic conditionally essential gene by decreasing the promoter activity, so that more than one functional copy of the gene would be advantageous to a deficient host cell. The integration of an amplification unit comprising such a low-level expression conditionally essential gene, into a host cell deficient for the same gene, reproducibly resulted in genomic duplications of the integrated amplification unit, comparable to what has been observed when using traditional amplifiable antibiotic markers.

In fact, this invention provides the means for controlling the level of gene expression, *i.e.*, copy-number, in a host cell. By choosing carefully the strength of the heterologous promoter expressing the conditionally essential marker gene in the amplification unit, the optimal copy-number of the amplification unit can be adjusted up or down, depending on the desired expression level of the gene of interest also comprised in the unit.

Accordingly, in a first aspect the invention relates to a bacterial host cell comprising at least two copies of an amplification unit in its genome, said amplification unit comprising:

i)   at least one copy of a gene of interest, and

ii) an expressible conditionally essential gene, wherein the conditionally essential gene is either promoterless or transcribed from a heterologous promoter having an activity substantially lower than the endogenous promoter of said conditionally essential gene, and

wherein the conditionally essential gene if not functional would render the cell auxotrophic for at least one specific substance or unable to utilize one or more specific sole carbon source.

In a second aspect, the invention relates to a method for producing a protein encoded by a gene of interest, comprising

a)   culturing a bacterial host cell comprising at least two duplicated copies of an amplification unit in its genome, the amplification unit comprising:

i)   at least one copy of the gene of interest, and

ii) an expressible conditionally essential gene, wherein the conditionally essential gene is either promoterless or transcribed from a heterologous promoter having an activity substantially lower than the endogenous promoter of said conditionally essential gene,

wherein the conditionally essential gene if not functional would render the cell auxotrophic for at least one specific substance or unable to utilize one or more specific sole carbon source; and

b)   recovering the protein.

In a final aspect, the invention also relates to a method for producing a bacterial cell comprising two or more amplified chromosomal copies of a gene of interest, the method comprising:

a) providing a bacterial cell comprising at least one copy of an amplification unit, the unit comprising:

i)   at least one copy of the gene of interest, and

ii) an expressible functional copy of a conditionally essential gene, which is either promoterless or transcribed from a heterologous promoter having an activity

3

substantially lower than the endogenous promoter of said conditionally essential gene,

wherein the conditionally essential gene if not functional would render the cell auxotrophic for at least one specific substance or unable to utilize one or more specific sole carbon

5   source;

b) cultivating the cell under conditions suitable for growth in a medium deficient of said at least one specific substance and/or with said one or more specific sole carbon source, thereby providing a growth advantage to a cell in which the amplification unit has been duplicated in the chromosome; and

10   c) selecting a cell wherein the amplification unit has been duplicated in the chromosome, whereby two or more amplified chromosomal copies of the gene of interest were produced.

It is envisioned that all the preferred embodiments of the cell of the invention that are shown herein would be suitable for use in the methods of the second and third aspects

15   of the invention

DEFINITIONS

In accordance with the present invention there may be employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art.

20   Such techniques are explained fully in the literature. See, e.g., Sambrook, Fritsch & Maniatis, Molecular Cloning: A Laboratory Manual, Second Edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (herein "Sambrook et al., 1989") DNA Cloning: A Practical Approach, Volumes I and II /D.N. Glover ed. 1985); Oligonucleotide Synthesis (M.J. Gait ed. 1984); Nucleic Acid Hybridization (B.D. Hames &

25   S.J. Higgins eds (1985)); Transcription And Translation (B.D. Hames & S.J. Higgins, eds. (1984)); Animal Cell Culture (R.I. Freshney, ed. (1986)); Immobilized Cells And Enzymes (IRL Press, (1986)); B. Perbal, A Practical Guide To Molecular Cloning (1984).

A "polynucleotide" is a single- or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. Polynucleotides include RNA and DNA,

30   and may be isolated from natural sources, synthesized in vitro, or prepared from a combination of natural and synthetic molecules.

A "nucleic acid molecule" or "nucleotide sequence" refers to the phosphate ester polymeric form of ribonucleosides (adenosine, guanosine, uridine or cytidine; "RNA molecules") or deoxyribonucleosides (deoxyadenosine, deoxyguanosine, deoxythymidine, or

35   deoxycytidine; "DNA molecules") in either single stranded form, or a double-stranded helix. Double stranded DNA-DNA, DNA-RNA and RNA-RNA helices are possible. The term nucleic acid molecule, and in particular DNA or RNA molecule, refers only to the primary and

secondary structure of the molecule, and does not limit it to any particular tertiary or quaternary forms. Thus, this term includes double-stranded DNA found, inter alia, in linear or circular DNA molecules (e.g., restriction fragments), plasmids, and chromosomes. In discussing the structure of particular double-stranded DNA molecules, sequences may be
5   described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the nontranscribed strand of DNA (i.e., the strand having a sequence homologous to the mRNA). A "recombinant DNA molecule" is a DNA molecule that has undergone a molecular biological manipulation.

A nucleic acid molecule is "hybridizable" to another nucleic acid molecule, such as
10   a cDNA, genomic DNA, or RNA, when a single stranded form of the nucleic acid molecule can anneal to the other nucleic acid molecule under the appropriate conditions of temperature and solution ionic strength (see Sambrook et al., supra). The conditions of temperature and ionic strength determine the "stringency" of the hybridization.

A DNA "coding sequence" or an "open reading frame (ORF)" is a double-stranded
15   DNA sequence which is transcribed and translated into a polypeptide in a cell in vitro or in vivo when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxyl) terminus. A coding sequence can include, but is not limited to, prokaryotic sequences, cDNA from eukaryotic mRNA, genomic DNA
20   sequences from eukaryotic (e.g., mammalian) DNA, and even synthetic DNA sequences. If the coding sequence is intended for expression in a eukaryotic cell, a polyadenylation signal and transcription termination sequence will usually be located 3' to the coding sequence.

An expression vector is a DNA molecule, linear or circular, that comprises a segment encoding a polypeptide of interest operably linked to additional segments that
25   provide for its transcription. Such additional segments may include promoter and terminator sequences, and optionally one or more origins of replication, one or more selectable markers, an enhancer, a polyadenylation signal, and the like. Expression vectors are generally derived from plasmid or viral DNA, or may contain elements of both.

Transcriptional and translational control sequences are DNA regulatory sequences,
30   such as promoters, enhancers, terminators, and the like, that provide for the expression of a coding sequence in a host cell. In eukaryotic cells, polyadenylation signals are control sequences.

A "secretory signal sequence" is a DNA sequence that encodes a polypeptide (a "secretory peptide" that, as a component of a larger polypeptide, directs the larger
35   polypeptide through a secretory pathway of a cell in which it is synthesized. The larger polypeptide is commonly cleaved to remove the secretory peptide during transit through the secretory pathway.

The term "promoter" is used herein for its art-recognized meaning to denote a portion of a gene containing DNA sequences that provide for the binding of RNA polymerase and initiation of transcription. Promoter sequences are commonly, but not always, found in the 5' non-coding regions of genes.

5   A chromosomal gene is rendered non-functional if the polypeptide that the gene encodes can no longer be expressed in a functional form. Such non-functionality of a gene can be induced by a wide variety of genetic manipulations as known in the art, some of which are described in Sambrook et al. vide supra. Partial deletions within the ORF of a gene will often render the gene non-functional, as will mutations.

10   The term "an expressible copy of a chromosomal gene" is used herein as meaning a copy of the ORF of a chromosomal gene, wherein the ORF can be expressed to produce a fully functional gene product. The expressible copy may not be transcribed from the native promoter of the chromosomal gene, it may instead be transcribed from a foreign or heterologous promoter, or it may indeed be promoterless and expressed only by

15   transcriptional read-through from a gene present upstream of the 5' end of the ORF. Transcriptional read-through is intended to have the same meaning here as the generally recognized meaning in the art.

"Operably linked", when referring to DNA segments, indicates that the segments are arranged so that they function in concert for their intended purposes, e.g. transcription

20   initiates in the promoter and proceeds through the coding segment to the terminator.

A coding sequence is "under the control" of transcriptional and translational control sequences in a cell when RNA polymerase transcribes the coding sequence into mRNA, which is then trans-RNA spliced and translated into the protein encoded by the coding sequence.

25   "Heterologous" DNA refers to DNA not naturally located in the cell, or in a chromosomal site of the cell. Preferably, the heterologous DNA includes a gene foreign to the cell.

As used herein the term "nucleic acid construct" is intended to indicate any nucleic acid molecule of cDNA, genomic DNA, synthetic DNA or RNA origin. The term "construct"

30   is intended to indicate a nucleic acid segment which may be single- or double-stranded, and which may be based on a complete or partial naturally occurring nucleotide sequence encoding a polypeptide of interest. The construct may optionally contain other nucleic acid segments.

The nucleic acid construct of the invention encoding the polypeptide of the invention

35   may suitably be of genomic or cDNA origin, for instance obtained by preparing a genomic or cDNA library and screening for DNA sequences coding for all or part of the polypeptide by

6

hybridization using synthetic oligonucleotide probes in accordance with standard techniques (cf. Sambrook et al., supra).

5   The nucleic acid construct of the invention encoding the polypeptide may also be prepared synthetically by established standard methods, e.g. the phosphoamidite method described by Beaucage and Caruthers, Tetrahedron Letters 22 (1981), 1859 - 1869, or the method described by Matthes et al., EMBO Journal 3 (1984), 801 - 805. According to the phosphoamidite method, oligonucleotides are synthesized, e.g. in an automatic DNA synthesizer, purified, annealed, ligated and cloned in suitable vectors.

10   Furthermore, the nucleic acid construct may be of mixed synthetic and genomic, mixed synthetic and cDNA or mixed genomic and cDNA origin prepared by ligating fragments of synthetic, genomic or cDNA origin (as appropriate), the fragments corresponding to various parts of the entire nucleic acid construct, in accordance with standard techniques. The nucleic acid construct may also be prepared by polymerase chain reaction using specific primers, for instance as described in US 4,683,202 or Saiki et al.,
15   Science 239 (1988), 487 – 491.

The term nucleic acid construct may be synonymous with the term "expression cassette" when the nucleic acid construct contains the control sequences necessary for expression of a coding sequence of the present invention

The term "control sequences" is defined herein to include all components which are
20   necessary or advantageous for expression of the coding sequence of the nucleic acid sequence. Each control sequence may be native or foreign to the nucleic acid sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader, a polyadenylation sequence, a propeptide sequence, a promoter, a signal sequence, and a transcription terminator. At a minimum, the control sequences include a promoter, and
25   transcriptional and translational stop signals. The control sequences may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the nucleic acid sequence encoding a polypeptide.

The control sequence may be an appropriate promoter sequence, a nucleic acid sequence which is recognized by a host cell for expression of the nucleic acid sequence.
30   The promoter sequence contains transcription and translation control sequences which mediate the expression of the polypeptide. The promoter may be any nucleic acid sequence which shows transcriptional activity in the host cell of choice and may be obtained from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell.

35   The control sequence may also be a suitable transcription terminator sequence, a sequence recognized by a host cell to terminate transcription. The terminator sequence is operably linked to the 3' terminus of the nucleic acid sequence encoding the polypeptide.

Any terminator which is functional in the host cell of choice may be used in the present invention.

The control sequence may also be a polyadenylation sequence, a sequence which is operably linked to the 3' terminus of the nucleic acid sequence and which, when transcribed, is recognized by the host cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence which is functional in the host cell of choice may be used in the present invention.

The control sequence may also be a signal peptide coding region, which codes for an amino acid sequence linked to the amino terminus of the polypeptide which can direct the expressed polypeptide into the cell's secretory pathway of the host cell. The 5' end of the coding sequence of the nucleic acid sequence may inherently contain a signal peptide coding region naturally linked in translation reading frame with the segment of the coding region which encodes the secreted polypeptide. Alternatively, the 5' end of the coding sequence may contain a signal peptide coding region which is foreign to that portion of the coding sequence which encodes the secreted polypeptide. A foreign signal peptide coding region may be required where the coding sequence does not normally contain a signal peptide coding region. Alternatively, the foreign signal peptide coding region may simply replace the natural signal peptide coding region in order to obtain enhanced secretion of the enzyme relative to the natural signal peptide coding region normally associated with the coding sequence. The signal peptide coding region may be obtained from a glucoamylase or an amylase gene from an *Aspergillus* species, a lipase or proteinase gene from a *Rhizomucor* species, the gene for the alpha-factor from *Saccharomyces cerevisiae*, an amylase or a protease gene from a *Bacillus* species, or the calf preprochymosin gene. However, any signal peptide coding region capable of directing the expressed polypeptide into the secretory pathway of a host cell of choice may be used in the present invention.

The control sequence may also be a propeptide coding region, which codes for an amino acid sequence positioned at the amino terminus of a polypeptide. The resultant polypeptide is known as a proenzyme or propolypeptide (or a zymogen in some cases). A propolypeptide is generally inactive and can be converted to mature active polypeptide by catalytic or autocatalytic cleavage of the propeptide from the propolypeptide. The propeptide coding region may be obtained from the *Bacillus subtilis* alkaline protease gene (aprE), the *Bacillus subtilis* neutral protease gene (nprT), the *Saccharomyces cerevisiae* alpha-factor gene, or the *Myceliophthora thermophilum* laccase gene (WO 95/33836).

It may also be desirable to add regulatory sequences which allow the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those which cause the expression of the gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory

8

compound. Regulatory systems in prokaryotic systems would include the lac, tac, and trp operator systems.   In yeast, the ADH2 system or GAL1 system may be used.   In filamentous fungi, the TAKA alpha-amylase promoter, *Aspergillus niger* glucoamylase promoter, and the *Aspergillus oryzae* glucoamylase promoter may be used as regulatory

5   sequences.   Other examples of regulatory sequences are those which allow for gene amplification.   In eukaryotic systems, these include the dihydrofolate reductase gene which is amplified in the presence of methotrexate, and the metallothionein genes which are amplified with heavy metals.   In these cases, the nucleic acid sequence encoding the polypeptide would be placed in tandem with the regulatory sequence.

10   Examples of suitable promoters for directing the transcription of the conditionally essential gene(s) of the present invention, especially in a bacterial host cell, are the promoters obtained from the *E. coli* lac operon, the *Streptomyces coelicolor* agarase gene (dagA), the *Bacillus subtilis* levansucrase gene (sacB), the *Bacillus subtilis* alkaline protease gene, the *Bacillus licheniformis* alpha-amylase gene (amyL), the *Bacillus stearothermophilus*

15   maltogenic amylase gene (amyM), the *Bacillus amyloliquefaciens* alpha-amylase gene (amyQ), the *Bacillus amyloliquefaciens* BAN amylase gene, the *Bacillus licheniformis* penicillinase gene (penP), the *Bacillus subtilis* xylA and xylB genes, and the prokaryotic beta-lactamase gene (Villa-Kamaroff et al., 1978, Proceedings of the National Academy of Sciences USA 75:3727-3731), as well as the tac promoter (DeBoer et al., 1983,

20   Proceedings of the National Academy of Sciences USA 80:21-25). Further promoters are described in "Useful proteins from recombinant bacteria" in Scientific American, 1980, 242:74-94; and in Sambrook et al., 1989, supra.

The term "auxotrophic" in the present context means that the auxotrophic cell requires at least one specific substance for growth and metabolism that the parental

25   organism was able to synthesize on its own. The term is used with respect to organisms, such as strains of bacteria, that can no longer synthesize the substance(s) because of mutational changes.

An effective signal peptide coding region for bacterial host cells is the signal peptide coding region obtained from the maltogenic amylase gene from *Bacillus* NCIB 11837, the

30   *Bacillus stearothermophilus* alpha-amylase gene, the *Bacillus licheniformis* subtilisin gene, the *Bacillus licheniformis* beta-lactamase gene, the *Bacillus stearothermophilus* neutral proteases genes (nprT, nprS, nprM), and the *Bacillus subtilis* PrsA gene. Further signal peptides are described by Simonen and Palva, 1993, Microbiological Reviews 57:109-137.

The present invention also relates to recombinant expression vectors comprising a

35   nucleic acid sequence of the present invention, a promoter, and transcriptional and translational stop signals. The various nucleic acid and control sequences described above may be joined together to produce a recombinant expression vector which may include one

or more convenient restriction sites to allow for insertion or substitution of the nucleic acid sequence encoding the polypeptide at such sites. Alternatively, the nucleic acid sequence of the present invention may be expressed by inserting the nucleic acid sequence or a nucleic acid construct comprising the sequence into an appropriate vector for expression.

5   In creating the expression vector, the coding sequence is located in the vector so that the coding sequence is operably linked with the appropriate control sequences for expression, and possibly secretion.

The recombinant expression vector may be any vector (e.g., a plasmid or virus) which can be conveniently subjected to recombinant DNA procedures and can bring about

10   the expression of the nucleic acid sequence. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vectors may be linear or closed circular plasmids. The vector may be an autonomously replicating vector, i.e., a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g., a plasmid, an extrachromosomal

15   element, a minichromosome, or an artificial chromosome. The vector may contain any means for assuring self-replication. Alternatively, the vector may be one which, when introduced into the host cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. The vector system may be a single vector or plasmid or two or more vectors or plasmids which together contain the total DNA to

20   be introduced into the genome of the host cell, or a transposon.

The vectors of the present invention preferably contain one or more selectable markers which permit easy selection of transformed cells. A selectable marker is a gene the product of which provides for biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophs, and the like.

25   Antibiotic selectable markers confer antibiotic resistance to such antibiotics as ampicillin, kanamycin, chloramphenicol, tetracycline, neomycin, hygromycin or methotrexate. Suitable markers for yeast host cells are ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3.

The vectors of the present invention preferably contain an element(s) that permits

30   stable integration of the vector, or of a smaller part of the vector, into the host cell genome or autonomous replication of the vector in the cell independent of the genome of the cell.

The vectors, or smaller parts of the vectors such as amplification units of the present invention, may be integrated into the host cell genome when introduced into a host cell. For chromosomal integration, the vector may rely on the nucleic acid sequence

35   encoding the polypeptide or any other element of the vector for stable integration of the vector into the genome by homologous or nonhomologous recombination.

Alternatively, the vector may contain additional nucleic acid sequences for directing integration by homologous recombination into the genome of the host cell. The additional nucleic acid sequences enable the vector to be integrated into the host cell genome at a precise location(s) in the chromosome(s). To increase the likelihood of integration at a precise location, the integrational elements should preferably contain a sufficient number of nucleic acids, such as 100 to 1,500 base pairs, preferably 400 to 1,500 base pairs, and most preferably 800 to 1,500 base pairs, which are highly homologous with the corresponding target sequence to enhance the probability of homologous recombination. The integrational elements may be any sequence that is homologous with the target sequence in the genome of the host cell. Furthermore, the integrational elements may be non-encoding or encoding nucleic acid sequences; specific examples of encoding sequences suitable for site-specific integration by homologous recombination are given in WO 02/00907 (Novozymes, Denmark), which is hereby incorporated by reference in its totality.

On the other hand, the vector may be integrated into the genome of the host cell by non-homologous recombination. These nucleic acid sequences may be any sequence that is homologous with a target sequence in the genome of the host cell, and, furthermore, may be non-encoding or encoding sequences. The copy number of a vector, an expression cassette, an amplification unit, a gene or indeed any defined nucleotide sequence is the number of identical copies that are present in a host cell at any time. A gene or another defined chromosomal nucleotide sequence may be present in one, two, or more copies on the chromosome. An autonomously replicating vector may be present in one, or several hundred copies per host cell.

An amplification unit of the invention is a nucleotide sequence that can integrate into the chromosome of a host cell, whereupon it can increase in number of chromosomally integrated copies by duplication of multiplication. The unit comprises an expression cassette as defined herein comprising at least one copy of a gene of interest and an expressable copy of a chromosomal gene, as defined herein, of the host cell. When the amplification unit is integrated into the chromosome of a host cell, it is defined as that particular region of the chromosome which is prone to being duplicated by homologous recombination between two directly repeated regions of DNA. The precise border of the amplification unit with respect to the flanking DNA is thus defined functionally, since the duplication process may indeed duplicate parts of the DNA which was introduced into the chromosome as well as parts of the endogenous chromosome itself, depending on the exact site of recombination within the repeated regions. This principle is illustrated in Janniére *et al.* (1985, Stable gene amplification in the chromosome of *Bacillus subtilis*. Gene, 40: 47-55), which is incorporated herein by reference.

For autonomous replication, the vector may further comprise an origin of replication enabling the vector to replicate autonomously in the host cell in question. Examples of bacterial origins of replication are the origins of replication of plasmids pBR322, pUC19, pACYC177, pACYC184, pUB110, pE194, pTA1060, and pAMbeta1. Examples of origin of replications for use in a yeast host cell are the 2 micron origin of replication, the combination of CEN6 and ARS4, and the combination of CEN3 and ARS1. The origin of replication may be one having a mutation which makes its functioning temperature-sensitive in the host cell (see, e.g., Ehrlich, 1978, Proceedings of the National Academy of Sciences USA 75:1433).

The present invention also relates to recombinant host cells, comprising a nucleic acid sequence of the invention, which are advantageously used in the recombinant production of the polypeptides. The term "host cell" encompasses any progeny of a parent cell which is not identical to the parent cell due to mutations that occur during replication.

The cell is preferably transformed with a vector comprising a nucleic acid sequence of the invention followed by integration of the vector into the host chromosome. "Transformation" means introducing a vector comprising a nucleic acid sequence of the present invention into a host cell so that the vector is maintained as a chromosomal integrant or as a self-replicating extra-chromosomal vector. Integration is generally considered to be an advantage as the nucleic acid sequence is more likely to be stably maintained in the cell. Integration of the vector into the host chromosome may occur by homologous or non-homologous recombination as described above.

The transformation of a bacterial host cell may, for instance, be effected by protoplast transformation (see, e.g., Chang and Cohen, 1979, Molecular General Genetics 168:111-115), by using competent cells (see, e.g., Young and Spizizin, 1961, Journal of Bacteriology 81:823-829, or Dubnar and Davidoff-Abelson, 1971, Journal of Molecular Biology 56:209-221), by electroporation (see, e.g., Shigekawa and Dower, 1988, Biotechniques 6:742-751), or by conjugation (see, e.g., Koehler and Thorne, 1987, Journal of Bacteriology 169:5771-5278).

The transformed or transfected host cells described above are cultured in a suitable nutrient medium under conditions permitting the expression of the desired polypeptide, after which the resulting polypeptide is recovered from the cells, or the culture broth.

The medium used to culture the cells may be any conventional medium suitable for growing the host cells, such as minimal or complex media containing appropriate supplements. Suitable media are available from commercial suppliers or may be prepared according to published recipes (e.g. in catalogues of the American Type Culture Collection). The media are prepared using procedures known in the art (see, e.g., references for bacteria and yeast; Bennett, J.W. and LaSure, L., editors, More Gene Manipulations in Fungi, Academic Press, CA, 1991).

If the polypeptide is secreted into the nutrient medium, the polypeptide can be recovered directly from the medium. If the polypeptide is not secreted, it is recovered from cell lysates. The polypeptide are recovered from the culture medium by conventional procedures including separating the host cells from the medium by centrifugation or filtration,

5   precipitating the proteinaceous components of the supernatant or filtrate by means of a salt, e.g. ammonium sulphate, purification by a variety of chromatographic procedures, e.g. ion exchange chromatography, gelfiltration chromatography, affinity chromatography, or the like, dependent on the type of polypeptide in question.

The polypeptides may be detected using methods known in the art that are specific

10   for the polypeptides. These detection methods may include use of specific antibodies, formation of an enzyme product, or disappearance of an enzyme substrate. For example, an enzyme assay may be used to determine the activity of the polypeptide.

The polypeptides of the present invention may be purified by a variety of procedures known in the art including, but not limited to, chromatography (e.g., ion

15   exchange, affinity, hydrophobic, chromatofocusing, and size exclusion), electrophoretic procedures (e.g., preparative isoelectric focusing (IEF), differential solubility (e.g., ammonium sulfate precipitation), or extraction (see, e.g., Protein Purification, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989).

DETAILED DESCRIPTION OF THE INVENTION

20   The first aspect of the invention relates to a bacterial host cell comprising at least two copies of an amplification unit in its genome, said amplification unit comprising:

   i) at least one copy of a gene of interest, and

   ii) an expressible conditionally essential gene, wherein the conditionally essential gene

25   is either promoterless or transcribed from a heterologous promoter having an activity substantially lower than the endogenous promoter of said conditionally essential gene, and

wherein the conditionally essential gene if not functional would render the cell auxotrophic for at least one specific substance or unable to utilize one or more specific sole carbon

30   source.

The choice of a host cell will to a large extent depend upon the gene encoding the polypeptide and its source. The host cell may be a unicellular microorganism, e.g., a prokaryote, or a non-unicellular microorganism, e.g., a eukaryote. Useful unicellular cells are bacterial cells such as gram positive bacteria including, but not limited to, a *Bacillus* cell,

35   e.g., *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus coagulans*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus stearothermophilus*, *Bacillus subtilis*, and *Bacillus thuringiensis*; or a

13

Streptomyces cell, e.g., *Streptomyces lividans* or *Streptomyces murinus*, or gram negative bacteria such as *E. coli* and *Pseudomonas sp.* In a preferred embodiment, the bacterial host cell is a *Bacillus lentus, Bacillus licheniformis, Bacillus stearothermophilus* or *Bacillus subtilis* cell. In one preferred embodiment, the bacterial host cell is a prokaryotic cell,

5   preferably a a Gram-positive prokaryotic cell, and more preferably the bacterial Gram positive cell is a species of the genus *Bacillus*, preferably selected from the group consisting of *Bacillus alkalophilus, Bacillus amyloliquefaciens, Bacillus brevis, Bacillus circulans, Bacillus coagulans, Bacillus lautus, Bacillus lentus, Bacillus licheniformis, Bacillus megaterium, Bacillus stearothermophilus, Bacillus subtilis, and Bacillus thuringiensis.*

10   As described above, chromosomal integration of a vector or a smaller part of a vector, such as an amplification unit of the invention, into the genome of the host cell can be achieved by a number of ways. A non-limiting example of integration by homologous recombination is shown herein.

A preferred embodiment of the invention relates to the cells of the invention, or the

15   methods of the invention, wherein the amplification unit further comprises a nucleotide sequence with a homology to a chromosomal nucleotide sequence of the host cell sufficient to effect chromosomal integration in the host cell of the amplification unit by homologous recombination, preferably the amplification unit further comprises a nucleotide sequence of at least 100 bp, preferably 200 bp, more preferably 300 bp, even more preferably 400 bp,

20   and most preferably at least 500 bp with an identity of at least 70%, preferably 80%, more preferably 90%, even more preferably 95%, and most preferably at least 98% identity to a chromosomal nucleotide sequence of the host cell.

In a non-limiting example integration into the chromosome of a host cell can be selected for by first rendering a conditionally essential host cell gene non-functional as

25   described elsewhere herein, thereby rendering the host cell selectable, then targetting the vector's integration by including on this a likewise non-functional copy of same host gene of a size that allows homologous recombination between the two different copies of the non-functional host genes in the genome of the host cell and on the integration vector, tailored so that such a recombination will restore a functional copy of the gene, thus leaving the host

30   cell selectable. Or the vector may simply comprise a functional copy of the conditionally essential gene, to select for integration anywhere in the genome.

A preferred embodiment of the invention relates to the cell of the invention, wherein a first amplification unit integrates into the host cell chromosome by homologous recombination with the partially deleted conditionally essential gene and renders the gene

35   functional.

A preferred embodiment of the invention relates to the cell of the invention, wherein the gene of interest encodes a polypeptide of interest, preferably the polypeptide is an

enzyme such as a protease; a cellulase; a lipase; a xylanase; a phospholipase; or preferably an amylase.

Another preferred embodiment of the invention relates to the cell of the invention, wherein the polypeptide is a hormone, a pro-hormone, a pre-pro-hormone, a small peptide,
5   a receptor, or a neuropeptide.

Still another preferred embodiment of the invention relates to the cell of the invention, wherein the gene of interest encodes an enzyme, preferably an amylolytic enzyme, a lipolytic enzyme, a proteolytic enzyme, a cellulytic enzyme, an oxidoreductase or a plant cell-wall degrading enzyme, and more preferably an enzyme with an activity selected
10   from the group consisting of aminopeptidase, amylase, amyloglucosidase, carbohydrase, carboxypeptidase, catalase, cellulase, chitinase, cutinase, cyclodextrin glycosyltransferase, deoxyribonuclease, esterase, galactosidase, beta-galactosidase, glucoamylase, glucose oxidase, glucosidase, haloperoxidase, hemicellulase, invertase, isomerase, laccase, ligase, lipase, lyase, mannosidase, oxidase, pectinase, peroxidase, phytase, phenoloxidase,
15   polyphenoloxidase, protease, ribonuclease, transferase, transglutaminase, or xylanase.

In a preferred embodiment, the invention relates to a cell, wherein the gene of interest encodes an antimicrobial peptide, preferably an anti-fungal peptide or an anti-bacterial peptide, or a peptide with biological activity in the human body, preferably a pharmaceutically active peptide, more preferably insulin/pro-insulin/pre-pro-insulin or
20   variants thereof, growth hormone or variants thereof, or blood clotting factor VII or VIII or variants thereof.

Conditionally essential genes are well-characterized in the literature, for instance genes that are required for a cell to synthesize one or more amino acids, where a non-functional gene encoding a polypeptide required for synthesis of an amino acid renders the
25   cell auxotrophic for that amino acid, and the cell can only grow if the amino acid is supplied to the growth medium. Restoration of the functionality of such a gene, or complementation by providing an exogenous functional copy of such a gene, allows the cell to synthesise the amino acid on its own, and it becomes selectable against a background of auxotrophic cells.

Consequently, a preferred embodiment of the invention relates to a cell of the first
30   aspect, wherein the conditionally essential chromosomal gene(s) of the host cell encodes one or more polypeptide(s) involved in amino acid synthesis, and the non-functionality of the endogenous versions of the gene(s) renders the cell auxotrophic for one or more amino acid(s), and wherein restoration of the functionality of the gene(s) renders the cell prototrophic for the amino acid(s).

35   *Bacillus subtilis metE* encodes a S-adenosyl-methionine synthetase, the *metE*/MetE sequences are available from EMBL:BS52812 (accession no. U52812) (Yocum,R.R.; Perkins,J.B.; Howitt,C.L.; Pero,J.; 1996. Cloning and characterization of the *metE* gene

encoding S-adenosylmethionine synthetase from *Bacillus subtilis*. J. Bacteriol. 178(15):4604).

The *leuB* gene encodes 3-isopropylmalate dehydrogenase, which catalyses the conversion of 3-carboxy-2-hydroxy-4-methylpentanoate to 3-carboxy-4-methyl-2-oxopentanoate. A *leuB*-deficient strain will be a leucine auxotroph.

The *lysA* gene encoding diaminopimelate decarboxylase, which catalyses the conversion of Meso-2,6-diaminoheptanedioate to L-lysine. A *lysA*-deficient strain will be a lysine auxotroph.

A preferred embodiment relates to a cell of the invention, wherein the conditionally essential gene encodes an enzyme from the biosynthetic pathway of an amino acid; preferably the conditionally essential gene encodes one or more polypeptide(s) involved in lysine, leucine or methionine synthesis, preferably the conditionally essential gene is homologous to the *lysA, leuB, metC*, or the *metE* gene from *Bacillus subtilis*, and more preferably the conditionally essential gene is the *lysA, leuB, metC*, or *metE* gene from *Bacillus licheniformis*; more preferably the conditionally essential gene is at least 75% identical, preferably 85% identical, more preferably 95% and most preferably at least 97% identical to the *lysA* sequence of *Bacillus licheniformis* shown in SEQ ID NO:48 of WO 02/00907 A1, the *leuB* sequence of *Bacillus licheniformis*, the *metC* sequence of *Bacillus licheniformis* shown in SEQ ID NO:42 of WO 02/00907 A1, or the *metE* sequence of *Bacillus subtilis* shown in positions 997 to 2199 of SEQ ID NO:16.

The *hemA* gene encodes glutamyl-tRNA reductase, which catalyses the synthesis of 5-amino leuvulinic acid. A *hemA*-deficient strain will have to be supplemented with 5-amino leuvulinic acid or haemin.

In another embodiment, the conditionally essential gene encodes a glutamyl-tRNA reductase, preferably the conditionally essential gene is homologous to the *hemA* gene from *Bacillus subtilis*, and more preferably the conditionally essential gene is the *hemA* gene from *Bacillus licheniformis*; preferably the conditionally essential gene is at least 75% identical, preferably 85% identical, more preferably 95% and most preferably at least 97% identical to the *hemA* sequence of *Bacillus licheniformis*.

The conditionally essential gene(s) may encode polypeptides involved in the utilization of specific carbon sources such as xylose, glucanate, glycerol, or arabinose, in which case the host cell is unable to grow in a minimal medium supplemented with only that specific carbon source when the gene(s) are non-functional.

A preferred embodiment of the invention relates to a cell of the invention, wherein the at least one conditionally essential chromosomal gene(s) is one or more genes that are required for the host cell to grow on minimal medium supplemented with only one specific main carbon-source.

A preferred embodiment relates to a cell of the invention, wherein the at least one conditionally essential gene encodes an enzyme required for xylose utilization, preferably the conditionally essential gene is homologous to the *xylA* gene from *Bacillus subtilis*, and more preferably the conditionally essential gene is homologous to a gene of the xylose

5   isomerase operon of *Bacillus licheniformis*, most preferably to the *xylA* gene of *Bacillus licheniformis*; preferably the conditionally essential gene encodes a xylose isomerase and is at least 75% identical, preferably 85% identical, more preferably 95% and most preferably at least 97% identical to the *xylA* gene of *Bacillus licheniformis*.

Another preferred embodiment relates to a cell of the invention, wherein the at least

10 one conditionally essential gene encodes an enzyme required for gluconate utilization, preferably the conditionally essential gene encodes a gluconate kinase (EC 2.7.1.12) or a gluconate permease, more preferably the gene is homologous to the *gntK* gene or the *gntP* gene from *Bacillus subtilis*, and most preferably the gene is the *gntK* or *gntP* gene from *Bacillus licheniformis*; preferably the conditionally essential gene encodes a gluconate

15 kinase (EC 2.7.1.12) or a gluconate permease or both and is at least 75% identical, preferably 85% identical, more preferably 95% and most preferably at least 97% identical to any of the *gntK* and *gntP* sequences of *Bacillus licheniformis*.

Still another preferred embodiment relates to a cell of the invention, wherein the conditionally essential gene encodes an enzyme required for glycerol utilization, preferably

20 the conditionally essential gene encodes a glycerol uptake facilitator (permease), a glycerol kinase, or a glycerol dehydrogenase, more preferably the conditionally essential gene is homologous to the *glpP, glpF, glpK,* or the *glpD* gene from *Bacillus subtilis*, and most preferably the conditionally essential gene comprises one or more of the *glpP, glpF, glpK,* and *glpD* genes from *Bacillus licheniformis* shown in SEQ ID NO:26 of published PCT

25 application WO 02/00907 A1 (Novozymes A/S) which is incorporated herein by reference in its totality; preferably the conditionally essential gene encodes a glycerol uptake facilitator (permease), a glycerol kinase, or a glycerol dehydrogenase, and is at least 75% identical, preferably 85% identical, more preferably 95% and most preferably at least 97% identical to any of the *glpP, glpF, glpK,* and *glpD* sequences of *Bacillus licheniformis* shown in SEQ ID

30 NO:26 of WO 02/00907 A1.

One more preferred embodiment relates to a cell of the invention, wherein the conditionally essential gene encodes an enzyme required for arabinose utilization, preferably an arabinose isomerase, more preferably the gene is homologous to the *araA* gene from *Bacillus subtilis*, and most preferably the gene is the *araA* gene from *Bacillus licheniformis*

35 shown in SEQ ID NO:38 of WO 02/00907 A1; preferably the conditionally essential gene encodes an arabinose isomerase, and is at least 75% identical, preferably 85% identical,

17

more preferably 95% and most preferably at least 97% identical to the *araA* sequence of *Bacillus licheniformis* shown in SEQ ID NO:38 of WO 02/00907 A1.

The amplification unit in the cell of the invention may also include an antibiotic marker gene. However, as it is preferred not to have marker genes in the chromosome, an alternative way of removing the marker gene must be employed. Specific restriction enzymes denoted resolvases excise portions of DNA if each portion is flanked on both sides by certain recognition sequences known as resolvase sites or *res*-sites; these resolvase enzymes are well-known in the art, see e.g. WO 96/23073 (Novo Nordisk A/S) which is included herein by reference.

A preferred embodiment relates to a cell of the invention, wherein the amplification unit further comprises an antibiotic selection marker, preferably the selection marker is flanked by resolvase sites or *res*-sites.

Subsequent to the action of the resolvase enzyme, the antibiotic restriction marker flanked by *res*-sites will have been excised from the chromosome of the cell, leaving only one copy of the *res*-site behind as testimony to the procedure.

Accordingly, a preferred embodiment relates to a cell of the invention, wherein the amplification unit further comprises a resolvase site or *res*-site.

As the present invention relies on a reduced transcription of the conditionally essential gene comprised in the amplification unit as compared to its wild-type transcription level, it may be an advantage to include one or more transcription terminators upstream of the gene in different reading frames, in order to avoid any unintentional read-through transcription from a gene further upstream in the chromosome from where the unit was integrated.

A preferred embodiment relates to a cell of the invention, wherein the conditionally essential gene comprised in the amplification unit has at least one transcription terminator located upstream of the gene.

Another way of reducing transcription of the conditionally essential gene is to express it from a heterologous or completely artificial promoter, which has a reduced activity as compared to the wild-type or endogenous promoter normally transcribing said gene. Preferably, the conditionally essential gene is transcribed from a heterologous promoter having an activity level, when compared with the endogenous promoter of the conditionally essential gene, which is reduced with a factor of 2, preferably 5, more preferably 10, even more preferably 50, and most preferably with a factor of 100.

Still another strategy could be to have a promoterless conditionally essential gene in the amplification unit, and then simply rely on what read-through transcription there might from any other gene(s) located upstream of the conditionally essential gene, before or after integration of the unit into the chromosome of the cell of the invention. Preferably, the

conditionally essential gene is promoterless; and more preferably the gene of interest is located upstream of the conditionally essential gene in the amplification unit, so that the two genes are co-directionally transcribed, whereby the conditionally essential gene is expressed by read-through transcription from the gene of interest.

5   A second aspect of the invention relates to a method for producing a protein encoded by a gene of interest, comprising

a)   culturing a bacterial host cell comprising at least two duplicated copies of an amplification unit in its genome, the amplification unit comprising:

i)   at least one copy of the gene of interest, and

10   ii)   an expressible conditionally essential gene, wherein the conditionally essential gene is either promoterless or transcribed from a heterologous promoter having an activity substantially lower than the endogenous promoter of said conditionally essential gene,

wherein the conditionally essential gene if not functional would render the cell auxotrophic
15   for at least one specific substance or unable to utilize one or more specific sole carbon source; and

b)   recovering the protein.

As already mentioned, any cell of the invention is envisioned to be suitable in the methods of the second aspect, in particular the preferred embodiments outlined in the
20   above.

A final aspect of the invention relates to methods for producing a bacterial cell comprising two or more amplified chromosomal copies of a gene of interest, the method comprising:

a) providing a bacterial cell comprising at least one copy of an amplification unit, the unit
25   comprising:

i)   at least one copy of the gene of interest, and

ii)   an expressible functional copy of a conditionally essential gene, which is either promoterless or transcribed from a heterologous promoter having an activity substantially lower than the endogenous promoter of said conditionally
30   essential gene,

wherein the conditionally essential gene if not functional would render the cell auxotrophic for at least one specific substance or unable to utilize one or more specific sole carbon source;

b) cultivating the cell under conditions suitable for growth in a medium deficient of said at
35   least one specific substance and/or with said one or more specific sole carbon source, thereby providing a growth advantage to a cell in which the amplification unit has been duplicated in the chromosome; and

19

c) selecting a cell wherein the amplification unit has been duplicated in the chromosome, whereby two or more amplified chromosomal copies of the gene of interest were produced.

5   Again, as already mentioned, the methods of the final aspect of the invention are envisioned as being suitable for producing any cell of the invention, in particular the preferred embodiments of said cell that are outlined in the above.

EXAMPLES

Strains and Donor Organisms

10   *Bacillus subtilis* PL1801. This strain is the *B.subtilis* DN1885 with disrupted *apr* and *npr* genes (Diderichsen, B., Wedsted, U., Hedegaard, L., Jensen, B. R., Sjøholm, C. (1990) Cloning of *aldB*, which encodes alpha-acetolactate decarboxylase, an exoenzyme from *Bacillus brevis*. J. Bacteriol., 172, 4315-4321).

      B.subtilis CLO46. This strain is a *B. subtilis* PL1801 where the *metE* gene is deleted
15   and replaced with the kanamycine (kan) resistance gene from pUB110 by use of the plasmid pCLO43.

      B.subtilis CLO49. This strain is the CLO46 strain where the kanamycine resistance gene is deleted.

      Competent cells were prepared and transformed as described by Yasbin, R.E.,
20   Wilson, G.A. and Young, F.E. (1975) Transformation and transfection in lysogenic strains of *Bacillus subtilis*: evidence for selective induction of prophage in competent cells. J. Bacteriol, 121:296-304.

Plasmids

25   pCLO43:

      This plasmid is a pBR322 derivative (Watson, N., 1988 Gene 70(2):399-403) essentially containing elements making the plasmid propagatable in *E. coli*, a ampicillin resistance gene, a gene conferring resistance to kanamycine, two flanking fragments from *B. subtilis metE* inserted upstream and downstream of the kanamycine resistance gene, two
30   direct repeats that signify the *res* site from pAMBeta1 (Janniere, L., 1996, Nucleic Acids Res. 24(17):3431-3436. This plasmid is used for deleting the *metE* gene in the *B. subtilis* strain PL1801.

Table 1. Plasmid pCLO43, 7311 bp
35

Position (bp)	Size (bp)	Element (bp)	Origin
1-973	973	Upstream *metE* seq.	*B. subtilis*

974-1010	37	Linker	*Synthetic*
10111-1184	174	*res* site from pAMbeta1	*E. faecalis*
1185-1190	6	Linker	*Synthetic*
1191-2159	969	pUB110 (Kan gene)	*S. aureus*
2160-2162	3	Linker	*Synthetic*
2163-2336	174	*res* site from pAMβ1	*E. faecalis*
2337-2357	21	Linker	*Synthetic*
2358-3870	1513	Downstream *metE* seq.	*B. subtilis*
3871-7311	3441	pBR322	*E. coli*

pCLO1154

This plasmid is a pBR322 derivative (Watson, N., 1988 Gene 70(2):399-403) containing elements making the plasmid propagatable in *E. coli*. The plasmid codes for the ampicillin resistance gene, the kanamycine resistance gene, the chloramphenicol resistance gene and the *lacZ* gene from *E. coli*. The *gfp* gene from *A. victoria* and the *metE* gene from *B. subtilis* are transcriptionally fused in the plasmid controlled by a promoter that can be exchanged with other promoters. This plasmid is used for integration and amplification studies in the *amyE* locus of CLO49. The primers for *metE* fragment PCR amplifications on chromosomal DNA isolated from PL1801 are as follows:

P52 (SEQ ID NO: 1): aataataaagatctggaggagaaacaatgacaacc
P53 (SEQ ID NO: 2): aaataataagatctaaattatactagctgtgtc

Table 2. Plasmid pCLO1154, 13135 bp.

Position (bp)	Size (bp)	Element (bp)	Origin
1-539	539	Upstream *amyE*	*B. subtilis*
540-2853	2314	*metE* gene	*B. subtilis*
2854-2891	38	Linker	Synthetic
2892-3605	714	*gfp* gene	*A. victoria*
3606-3739	134	Promoter - *alr*	*B. subtilis*
3740-3785	46	Linker	Synthetic
3786-4821	1036	pC194 (*cat* gene)	*S. aureus*

4822-5008	187	part of *tetC* gene	*E. coli*
5009-5106	98	Promoter	Synthetic
5107-5111	6	Linker	Synthetic
5112-8224	3113	*spoVG-lacZ* fusion	*B. subtilis & E. coli*
8226-8314	89	part of *tetC* gene	*E. coli*
8315-9657	1343	Downstream *amyE*	*B. subtilis*
9658-9845	188	Linker	Synthetic
9846-11117	1272	pUB110 (*neo* gene)	*S. aureus*
11118-11184	67	Linker	Synthetic
11185-11277	93	Tn5 fragment	*E. coli*
11278-11281	4	Linker	Synthetic
11282-13119	1838	pBR322 (*bla* gene)	*E. coli*
13120-13129	10	Linker	Synthetic

Propagation of PL1801 strain for LacZ activity determination

The *B. subtilis* strain PL1801 was propagated in liquid medium TY. After 10 generations of incubation at 37°C and 300 rpm, the cells were harvested, and cells were dis-
5   .rupted by sonic or lysozyme treatment.

General molecular biology methods

Unless otherwise mentioned the DNA manipulations and transformations were performed using standard methods of molecular biology (Sambrook et al. (1989) Molecular
10   cloning: A laboratory manual, Cold Spring Harbor lab., Cold Spring Harbor, NY; Ausubel, F. M. et al. (eds.) "Current protocols in Molecular Biology". John Wiley and Sons, 1995; Harwood, C. R., and Cutting, S. M. (eds.) "Molecular Biological Methods for Bacillus". John Wiley and Sons, 1990).

Enzymes for DNA manipulations were used according to the specifications of the
15   suppliers (e.g. restriction endonucleases, ligases etc. are obtainable from New England Bio-labs, Inc.).

Media

TY: (as described in Ausubel, F. M. et al. (eds.) "Current protocols in Molecular
20   Biology". John Wiley and Sons, 1995). LB agar (as described in Ausubel, F. M. et al. (eds.) "Current protocols in Molecular Biology". John Wiley and Sons, 1995).

Minimal TSS agar: As described in Fouet A. and Sonenshein, A. L. (1990) A Target for Carbon Source-Dependant Negative Regulation of the citB Promoter of Bacillus subtilis. J. Bacteriol., 172, 835-844. For plates, 2% agar was added and for methionine auxotropy determination the plates were supplemented with 50 microg/ml methionine.

5

Assay for beta-galactosidase activity

Beta-galactosidase activity was determined by a method using ortho-nitrophenyl-beta-D-galactopyranoside as enzymatic substrate. Under a specified set of conditions (temp., pH, reaction time, buffer conditions) a given amount of beta-galactosidase will

10   degrade a cer-tain amount of substrate and a yellow colour will be produced. The colour intensity is meas-ured at 420 nm. The measured absorbance is directly proportional to the activity of the beta-galactosidase in question under a given set of conditions.

15   **Deletion of *metE* in *B. subtilis***

A plasmid was constructed for the purpose of deleting the *metE* gene in *B. subtilis*. Two flanking sequences upstream and downstream of the *galE* gene were amplified by PCR and fused by PCR on each side of a kanamycine (Kana) marker. This fragment was ligated in plasmid pBR322.

20

Upstream metE fragment:

P42 (SEQ ID NO: 3): attttataggatcccgctgattcattttcttctgcgaac

P43 (SEQ ID NO: 4): gaattccatcgcactggacgacattttcaggtcgattctcggaaatcc

25   *Downstream metE fragment:*

P44 (SEQ ID NO: 5): cccgaggcctttcaggcccgcaaacaatatggttgaagccgcaaaacagg

P45 (SEQ ID NO: 6): ataataatggtaccatattgatgtgacacttgaagttgc

The resulting plasmid pCLO43 (SEQ ID NO: 7) was linearised and transferred to *B.*
30   *subtilis* PL1801 and plated on LBPG media with 10 µg/ml kanamycine, which left the Kan marker in place of the *metE* gene.

A *metE* deletion strain designated CLO46 was tested on minimal media without mehionine. The original *B. subtilis* PL1801 (*metE⁺*) strain showed fine growth on these plates while the *metE⁻* strain CLO46 showed no growth even after several days of
35   incubation. On control minimal plates supplemented with 50 µg/ml methionine, both strains grew. The reported auxotrophic phenotype on a *metE⁻* strain is therefore confirmed.

The Kan marker located in the *metE* locus of CLO46 was flanked by resolvase recognition sites *(res)*, which allow a specific excision reaction in the presence of a resolvase. In order to remove the Kan marker from the chromosome, CLO46 was transformed with pWT, which is a temperature sensitive plasmid that comprises a gene
5   coding for resolvase and an erythromycine (Erm) resistance marker. Transformants were selected on plates with 5microg/ml Erm. They were tested for loss of the Kan marker and further re-streaked twice on plates with no antibiotics at 50°C to cure the strains of the pWT plasmid. Selected clones were screened for loss of Erm resistance and Kan resistance and were designated CLO49 (PL1801, *metE*; no antibiotic markers).

10

Amplification plasmids

An amplification plasmid was made having a transcriptional unit concisting of the *gfp* gene and the *metE* gene with a cloning site in front of the two genes, wherein a promoter could be cloned (pCLO1154, SEQ ID NO: 8). The *lacZ* reporter gene was also present on
15   the plasmid expressed from a promoter separate from the promoter in front of the *metE* gene. Flanking these two transcriptional units was framgments from the *amyE* locus in *B. subtilis.*

Promoters with varying promoter activity were cloned in front of the *gfp-metE* transcriptional unit in the *EcoR*I and *Hind*III sites. The promoter activities spanned from 30 to
20   519 arbitrary units. See table 3.

Promoter	Activity / Units	Sequence
Pr30	30	(SEQ ID NO: 9)
Pr43	43	(SEQ ID NO: 10)
Pr119	119	(SEQ ID NO: 11)
Pr164	164	(SEQ ID NO: 12)
Pr342	342	(SEQ ID NO: 13)
Pr409	409	(SEQ ID NO: 14)
Pr519	519	(SEQ ID NO: 15)

Table 3: The table shows the promoters used in the amplification experiment and the sequence is given.

25

Amplification experiments

The resulting amplification plasmids were introduced by transformation into CLO49 (*metE*) and plated on solid LB media supplemented with 6 microg/ml chloramphenicol. Transformants were screened for resistance to kanamycine.

5   Transformants being sensitive to kanamycine would have integrated part of the amplification plasmid at the *amyE* locus including the *lacZ* reporter gene and the *gfp-metE* operon. Those transformans would have only one copy of the genes present and they cannot be amplified.

Transformants being resistant to kanamycine would have the whole amplification plasmid integrated at the *amyE* locus and amplification would be possible.

10   Both types of transformants were plated on solid minimal TSS media without methionine. Several colonies were obtained from the transformants having the whole plasmid integrated at the *amyE* locus, whereas the transformants that had only part of the plasmid integrated showed no growth on minimal medium. This indicated that even with the strongest promoter, one copy of the *metE* gene did not express sufficient MetE protein to

15   complement the methionine auxotrophy of the strain. However, amplification of the *metE* gene did result in growth of the strain.

Colonies were picked from the amplification step a long with colonies that had only one copy of the *metE* gene integrated in the chromosome. They were all grown in liquid LB and harvested in the exponential growth phase followed by measurement of □-

20   galactosidase activity. The following table gives the results from the evaluation of the amplification outcomes.

A few clones show irregular enzyme activities, which can be explained by up-mutations in the promoters.

Promoter Strength	Strain	Units	Copies
30	1 gene copy	105	1,0
	Amplification	1361	12,4
	Amplification	218	2,0
43	1 gene copy	101	0,9
	Amplification	1467	13,4
	Amplification	1460	13,3
119	1 gene copy	113	1,0
	Amplification	1055	9,6
	Amplification	1075	9,8
164	1 gene copy	102	0,9
	Amplification	881	8,0
	Amplification	855	7,8
342	1 gene copy	134	1,2
	Amplification	606	5,5
409	1 gene copy	105	1,0

	Amplification	533	4,9
	Amplification	493	4,5
519	1 gene copy	105	1,0
	Amplification	544	5,0
	Amplification	114	1,0

Table 4: The table shows the results from the amplification trials and the β-galactosidase activity measured in all strains after growth in LB lipuid media. The enzyme activities have been converted to the gene copy number of the reporter gene based on the enzyme activities.

The results summarized herein show that it is indeed possible to increase the copy number of a chromosomally integrated expression cassette holding a weakly expressed *metE* gene by growing the strain on minimal medium without methionine The amplification potential >10 copies (up to 25 copies have been observed), as judged from the enzyme activities is very similar to what can be achieved by the traditional kanamycine antibiotic selection/amplification.